



---

Year: 2001

---

## **Biochemical characterization of uracil processing activities in the hyperthermophilic archaeon *Pyrobaculum aerophilum*.**

Sartori, Alessandro A ; Schär, P ; Fitz-Gibbon, S ; Miller, J H ; Jiricny, J

**Abstract:** Deamination of cytosine to uracil and 5-methylcytosine to thymine represents a major mutagenic threat particularly at high temperatures. In double-stranded DNA, these spontaneous hydrolytic reactions give rise to G.U and G.T mispairs, respectively, that must be restored to G.C pairs prior to the next round of DNA replication; if left unrepaired, 50% of progeny DNA would acquire G.C → A.T transition mutations. The genome of the hyperthermophilic archaeon *Pyrobaculum aerophilum* has been recently shown to encode a protein, Pa-MIG, a member of the endonuclease III family, capable of processing both G.U and G.T mispairs. We now show that this latter activity is undetectable in crude extracts of *P. aerophilum*. However, uracil residues in G.U mispairs, in A.U pairs, and in single-stranded DNA were efficiently removed in these extracts. These activities were assigned to a approximately 22-kDa polypeptide named Pa-UDG (*P. aerophilum* uracil-DNA glycosylase). The recombinant Pa-UDG protein is highly thermostable and displays a considerable degree of homology to the recently described uracil-DNA glycosylases from *Archaeoglobus fulgidus* and *Thermotoga maritima*. Interestingly, neither Pa-MIG nor Pa-UDG was inhibited by UGI, a generic inhibitor of the UNG family of uracil glycosylases. Yet a small fraction of the total uracil processing activity present in crude extracts of *P. aerophilum* was inhibited by this peptide. This implies that the hyperthermophilic archaeon possesses at least a three-pronged defense against the mutagenic threat of hydrolytic deamination of cytosines in its genomic DNA.

DOI: <https://doi.org/10.1074/jbc.M102985200>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-34472>

Journal Article

Accepted Version

Originally published at:

Sartori, Alessandro A; Schär, P; Fitz-Gibbon, S; Miller, J H; Jiricny, J (2001). Biochemical characterization of uracil processing activities in the hyperthermophilic archaeon *Pyrobaculum aerophilum*. *Journal of Biological Chemistry*, 276(32):29979-29986.

DOI: <https://doi.org/10.1074/jbc.M102985200>

**DNA Replication Repair and  
Recombination:  
Biochemical characterization of uracil  
processing activities in the  
hyperthermophilic archaeon pyrobaculum  
aerophilum**

Alessandro A. Sartori, Primo Schar, Sorel  
Fitz-Gibbon, Jeffrey H. Miller and Josef  
Jiricny  
*J. Biol. Chem.* published online June 8, 2001

---

Access the most updated version of this article at doi: [10.1074/jbc.M102985200](https://doi.org/10.1074/jbc.M102985200)

Find articles, minireviews, Reflections and Classics on similar topics on the [JBC Affinity Sites](#).

Alerts:

- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at  
<http://www.jbc.org/content/early/2001/06/08/jbc.M102985200.citation.full.html#ref-list-1>

**Classification:** Enzymology

**Biochemical characterization of uracil processing activities in the hyperthermophilic archaeon *Pyrobaculum aerophilum*.**

**Alessandro A. Sartori<sup>1</sup>, Primo Schär<sup>1</sup>, Sorel Fitz-Gibbon<sup>2</sup>, Jeffrey H. Miller<sup>2</sup> and Josef Jiricny<sup>1\*</sup>**

**Institute of Medical Radiobiology  
of the University of Zürich and the Paul Scherrer-Institute  
August Forel-Strasse 7  
CH-8008 Zürich  
Switzerland<sup>1</sup>**

and

**Department of Microbiology and Molecular Genetics  
and the Molecular Biology Institute  
University of California,  
Los Angeles, CA 90095<sup>2</sup>**

**Running Title:** Uracil glycosylases in *Pyrobaculum aerophilum*

\*Communicating author

Tel.: +41-1-634 8910

Fax.: +41-1-634 8904

E-mail: jiricny@imr.unizh.ch

**ABSTRACT**

Deamination of cytosine to uracil and 5-methylcytosine to thymine represents a major mutagenic threat, particularly at high temperatures. In double-stranded DNA, these spontaneous hydrolytic reactions give rise to G•U and G•T mispairs, respectively, which must be restored to G•C pairs prior to the next round of DNA replication; left unrepaired, 50% of progeny DNA would acquire G•C → A•T transition mutations. The genome of the hyperthermophilic archaeon *Pyrobaculum aerophilum* was recently shown to encode a protein, *Pa-MIG*, a member of the endonuclease III family, capable of processing both G•U and G•T mispairs. We now show that this latter activity is undetectable in crude extracts of *P. aerophilum*. However, uracil residues in G•U mispairs, A•U pairs and in single-stranded DNA are efficiently removed in these extracts. These activities could be assigned to a ~22 kDa polypeptide named *Pa-UDG*. The recombinant *Pa-UDG* protein is highly thermostable and displays a considerable degree of homology to the recently-described uracil DNA glycosylases from *Archaeoglobus fulgidus* and *Thermotoga maritima*. Interestingly, neither *Pa-MIG*, nor *Pa-UDG* are inhibited by UGI, a generic inhibitor of the UNG family of uracil glycosylases. Yet, a small fraction of the total uracil processing activity present in crude extracts of *P. aerophilum* could be inhibited by this peptide. This implies that this hyperthermophilic archaeon possesses at least a three-pronged defense against the mutagenic threat of hydrolytic deamination of cytosines in its genomic DNA.

## INTRODUCTION

The integrity of genomic DNA of all organisms is under constant attrition by a variety of chemical and physical agents. A particular threat is posed by hydrolytic damage to DNA bases that carry exocyclic amino groups. The most frequent of these is the deamination of cytosine (1), which results in the formation of uracil, and therefore of G•U mismatches in double-stranded DNA. As uracil will base pair with adenine during the next round of DNA replication, cytosine deamination will result in a G•C  $\rightarrow$  A•T transition mutation if unrepaired. Because cytosine hydrolysis is greatly enhanced at high temperatures (1), hyperthermophilic organisms might be expected to be more susceptible to this type of mutations. However, this does not appear to be the case, as the reported spontaneous mutation rates in thermophilic bacteria are similar to those observed in *E. coli* (2). Thus, hyperthermophilic organisms are apparently more efficient at repairing hydrolytic and oxidative damage to DNA bases (3). We decided to study how uracil is processed in the hyperthermophilic archaeon *Pyrobaculum aerophilum* (4), which has an optimal growth temperature of 100°C.

In most organisms, uracil is removed from DNA by uracil-DNA glycosylases (UDG) that are encoded by the *UNG* gene family (5). These highly-efficient enzymes remove the modified base from all substrates, be they G•U mismatches, A•U pairs or single stranded. Interestingly, some organisms including *Drosophila melanogaster* appear to lack UNG gene homologues. Several years ago, we identified a novel mismatch-specific DNA-glycosylase, TDG, which removes uracil and thymine from G•U and G•T mispairs, respectively (6-8). Although this enzyme is substantially less efficient than UDG *in vitro*, it appears to be widespread in nature, at least as far as its activity on G•U mispairs is concerned. *E. coli* contains the *mug* gene, which encodes a TDG homologue that can process G•U substrates, but not G•T mispairs (9). Eukaryotic TDG homologues are found also in *Drosophila melanogaster* and *Schizosaccharomyces pombe* (U. Hardeland, M. Bentele, J. Jiricny and P. Schär, manuscript in preparation). Due to their strong preference for G•U mispairs over A•U pairs or uracil in single-stranded DNA, these enzymes have been assigned primarily an antimutagenic role in DNA metabolism (9). Surprisingly, homology searches of the genome sequences available for several hyperthermophilic bacteria revealed neither UNG nor Mug homologues (10,11). However, Sandigursky and Franklin have recently identified uracil DNA-glycosylases encoded in the genomes of the thermophiles *Thermotoga maritima* (12) and *Archaeoglobus fulgidus* (13), which appear to be distantly related to the Mug class of enzymes. Interestingly, in contrast to the mismatch-specific Mug protein, these enzymes can remove uracil from DNA containing either A•U pairs or G•U mispairs, as well as from single-stranded DNA substrates.

A recent report by Miller and colleagues (14) described the characterization of a *P. aerophilum* DNA glycosylase that removes uracil and thymine from G•U and G•T mispairs,

respectively. This enzyme was shown to be a homologue of the *Methanobacterium thermoautotrophicum* MIG glycosylase, which belongs to the endonuclease III class of enzymes that normally process saturated thymine residues such as thymine glycol and dihydrothymine (5).

We set out to examine cell free extracts of the hyperthermophilic archaeon *P. aerophilum*, in order to characterize the uracil processing activities present. We now report that uracil containing DNA in these extracts is processed predominantly by a thermostable UDG homologue, *Pa*-UDG, which is most closely related to the *T. maritima* and *A. fulgidus* enzymes. Interestingly, while the activity of *Pa*-MIG was undetectable in the extracts, we obtained first evidence of the existence of a third uracil-processing activity, which appears to be, unlike *Pa*-MIG and *Pa*-UDG, susceptible to inhibition by the UGI peptide, a generic inhibitor of the UNG-type uracil glycosylases.

## EXPERIMENTAL PROCEDURES

### *Reagents and Oligonucleotides*

All oligonucleotides were synthesized by Microsynth (Switzerland). The substrate oligonucleotides were purified by PAGE. Restriction enzymes and uracil DNA glycosylase inhibitor (UGI) were supplied by New England BioLabs (Beverly, MA). All other chemicals and reagents used were purchased from Sigma, Boehringer Mannheim, Amresco, Epicentre Technologies or Merck, and were of analytical grade purity.

### *Bacterial strains, media and plasmids*

The *E. coli* strain DH5 $\alpha$  was used for all cloning experiments and for plasmid amplifications, and the strain BL21(DE3) was used for protein expression (15). The plasmid pET28-*paudg* is a derivative of pET28c(+) (Novagen) and was used for bacterial expression of an N-terminal 6Histidine-tagged *Pa*-UDG under the control of the Lac operator and the T7 RNA polymerase. LB- and SOC-media were prepared as described in Sambrook et al. 1989 (15).

### *PCR and sequencing primers*

The primers used for the PCR-amplification of *P. aerophilum* genomic DNA containing *paudg* were: *paudg*3888s: 5'GATCCATATGGCTAGCGATTTGCAAAAGCTTCATGAG<sup>3'</sup> and *paudg*4434as: 5'GTACGGATCCTCACTTAGACGGGTCTAAGAAG<sup>3'</sup>, carrying *Nhe*I and *Bam*HI restriction sites, respectively (bold), which were used for subsequent cloning of the amplified fragment into the pET-28c(+) vector. *paudg* specific sequencing primers were: *paudg*3807s, 5'ACGCCTTCTCTACAGC<sup>3'</sup>; *paudg*4183s, 5'CCCGTACCTCATAACAAC<sup>3'</sup> and

*paudg4167as*, 5'GTTGTATGAGGTACGGG<sup>3'</sup>. A T7 promoter-specific primer was used for sequencing across the cloning junctions in the pET28c(+) vector.

#### *P. aerophilum* extracts and purified proteins

*P. aerophilum* cultures were grown in the laboratory of Jeffrey Miller (University of California, Los Angeles, USA). The cell free extracts were prepared by Mahmud Shivji (ICRF, Clare Hall Laboratories, London, UK), according to the protocol described previously (16) in a buffer containing 50 mM Tris•HCl, pH 8.0; 15 mM MgCl<sub>2</sub>; 1 mM EDTA and 1 mM DTT. Aliquots of the extract were stored at -80°C.

The purified recombinant *P. aerophilum* mismatch-specific DNA-glycosylase *Pa*-MIG (14) was kindly provided by Hanjing Yang (University of California, Los Angeles, USA). The enzyme was stored at -80°C in a buffer containing 50 mM Tris•HCl, pH 7.6; 1 mM EDTA; 1 mM DTT; 30 mM NaCl and 50% glycerol.

#### *Sequence Analyses and Cloning of the P. aerophilum* UDG Gene (*paudg*)

The candidate protein coding region, PAE0651, was identified in the recently-completed genome sequence of *P. aerophilum* [(11); S. Fitz-Gibbon, personal communication]. Analyses of the *P. aerophilum* genome were performed with the Genetics Computer Group program package, Version 10, 1999 (17), for other database searches and comparisons, we used the BLAST, FASTA, ENTREZ and CLUSTALW services provided at the EMBL web page (<http://www.ebi.ac.uk>).

The PAE0651 open reading frame was amplified by PCR using the primers *paudg3888s* and *paudg4434as* containing *Nhe*I and *Bam*HI restriction sites, respectively, and ligated into the corresponding restriction sites in the bacterial expression vector pET28c(+). The DNA sequence of the resulting plasmid (pET28-*paudg*) was confirmed by cycle sequencing using the primers *paudg4183s*, *paudg 4167as* and the T7 promoter primer. This cloning strategy places the PAE0651 coding sequence downstream of a vector-derived sequence encoding an N-terminal Histidine<sup>6</sup>-tag followed by a thrombin cleavage site. The N-terminal amino acid sequence of the resulting His<sup>6</sup>-*Pa*-UDG fusion protein is therefore

MGSSHHHHHSSGLVPRGSHMAS**DLQKL**.... (bold residues indicate the authentic *Pa*-UDG N-terminus lacking the initiator methionine; the thrombin cleavage site is underlined).

#### *Expression and Purification of Pa*-UDG

The pET28-*paudg* expression construct was transformed into *E. coli* BL21(DE3) cells by electroporation. Following incubation at 30°C overnight on selective LB-plates supplemented with 2% D-glucose, single colonies were picked and used to inoculate 20 ml of LB-medium containing 50 µg/ml kanamycin and 2% D-glucose. Following an overnight incubation at



30°C, 20 ml of the culture were then used to inoculate 1 liter of LB-medium containing 50 µg/ml kanamycin, and the culture was grown at 30°C up to OD<sub>600</sub> of 0.7-0.9. The expression of *Pa*-UDG was then induced with 1mM isopropylthio-β-D-galactopyranoside (IPTG, Sigma). After 5 h of incubation at 30°C, the cells were spun down at 8000 rpm and 4°C for 10 min in a Sorvall SLA-3000 rotor.

The cell pellet was resuspended in 3 ml/mg sonication buffer (50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 10% glycerol, 1 mM imidazole, 0.25% Tween-20, 10 mM β-mercaptoethanol and 1 mM phenylmethylsulfonyl fluoride). The cells were lysed by sonication on ice (25 × 5 s bursts with intermittent chilling for 10 s). The sonicate was clarified by centrifugation at 15000 rpm at 4°C for 30 min in a Sorvall SS34 rotor. The supernatant (soluble fraction) was removed and incubated with gentle shaking for 1 h at 4°C with 2 ml Ni-NTA agarose (QIAGEN), pre-equilibrated in sonication buffer. The suspension was then packed into a disposable column, from which the unbound proteins were eluted with sonication buffer containing increasing concentrations of imidazole (1 × 10 column-volumes (cv) 5 mM imidazole, 2 × 2.5 cv 60 mM imidazole). The histidine-tagged *Pa*-UDG protein was eluted with 3 × 1 cv of sonication buffer containing 250 mM imidazole. The 250 mM fractions were pooled and dialyzed overnight at 4°C against 2 liters of binding buffer (50 mM sodium phosphate, pH 8.0, 10 mM NaCl, 10 % glycerol, 10 mM β-mercaptoethanol). The dialyzed fraction was loaded onto a 1-ml HiTrap SP FPLC column (Amersham Pharmacia Biotech) and the column was washed with 10 ml of binding buffer. The column was then eluted with a 20 ml linear gradient of 10-600 mM NaCl. The nearly-homogenous *Pa*-UDG protein eluted as a major peak in fractions containing 150-200 mM NaCl. These fractions were pooled and dialyzed against storage buffer (50 mM sodium phosphate, pH 8.0, 50 mM NaCl, 10 % glycerol, 10 mM β-mercaptoethanol). The concentration was estimated to be ~1 mg/ml, as determined by the method of Bradford (18). The protein was stored in small aliquots at -80°C. This preparation was assayed for UDG activity.

### *Glycosylase Assays*

The uracil glycosylase activity of *P. aerophilum* extracts or of the purified enzymes was monitored by means of a standardized “nicking assay”, as described by Hudepohl *et al.* (16). In this assay, removal of uracil or thymine from a fluorescently-labeled synthetic 60-mer oligonucleotide DNA substrate generates an abasic site, which can be subsequently cleaved by



treatment with hot alkali or AP-endonuclease. This reaction generates a 23-mer fragment that can be resolved from the uncleaved 60-mer by denaturing polyacrylamide gel electrophoresis. A 60-mer homoduplex with a G•C base pair at position 23 served as a control. This substrate can be cleaved with the restriction endonuclease *AccI* to generate the diagnostic 23-mer marker fragment shown in Figures 1B, 3 and 6A.

The assays were set up in 20  $\mu$ l reaction mixtures containing 50 mM Tris•HCl pH 8.0, 1 mM DTT, 50mM EDTA, 10 mM KCl, 1pmol of labeled DNA and 1 to 25  $\mu$ g of extract from *P. aerophilum*, or 0.1 to 10 pmol of the purified proteins. Incubation conditions varied as indicated in the text. Assays involving protein extracts were subjected to digestion with 1  $\mu$ l of proteinase K (10 mg/ml) for 30 min at 37°C, those with purified proteins (*Pa*-MIG or *Pa*-UDG) to hot alkaline treatment by the addition of NaOH to 100 mM and incubation for 10 min at 99°C. Then 0.5  $\mu$ l carrier tRNA (10 mg/ml), 1/10 volume of 3 M sodium acetate pH 5.2 and three volumes of ice cold absolute ethanol were added, and the samples were chilled for 1 h at -20°C. DNA precipitates were then collected by centrifugation for 15 min at 14000 rpm and 4°C, washed in 180  $\mu$ l 70% ethanol and dried for 2 min in a Speed-Vac. The pellet was dissolved in 10  $\mu$ l formamide buffer (90% deionized formamide, 1× TBE without dye markers), heated for 5 min at 95°C and immediately chilled on ice. Finally, the reaction products were separated on denaturing 15% polyacrylamide gels (pre-run for 20 min at 450 Volts) by running in 1× TBE for 8 min at 450 V and then at 300 V until the bromophenol blue dye (loaded in an adjacent lane) migrated to the bottom of the gel. The bands were visualized on a FluoroImager (Storm 860, Molecular Dynamics) and the signals were quantified using the ImageQuant software (Version 1.2, Molecular Dynamics).

## RESULTS

### *Uracil-processing activities in P. aerophilum extracts*

Spontaneous hydrolytic deamination of cytosines in DNA is greatly enhanced at high temperatures (1) and generates uracil bases mispaired with guanines. Since the optimal growth temperature of *P. aerophilum* is 100°C, but its spontaneous mutation rate is no higher than that of *E. coli* (2), we anticipated that this organism should possess an efficient base excision repair mechanism, capable of rapid restoration of the original G•C base pairs. We first decided to examine the removal of uracil from DNA in crude extracts of *P. aerophilum*. This was measured as the efficiency of uracil-directed DNA strand incision upon incubation of different DNA substrates with the protein extract. As shown in Figure 1, the *P. aerophilum* extracts were

capable of specific removal of uracil from the double-stranded oligonucleotide substrate containing a single G•U mispair (Fig. 1A, lanes 3 and 4), as well as from an A•U base pair and from a single-stranded oligonucleotide (Fig. 1B) albeit with lower efficiency. In all cases, the reaction products migrated at the position of the 23-mer marker which is indicative of a concerted action of uracil DNA glycosylase and AP-endonuclease activities. The G•U processing reaction was reduced by approximately 10% after the addition of an excess of UGI (Fig. 1A, lane 4), a small polypeptide which specifically inhibits all prokaryotic and eukaryotic UDG homologues encoded by the *UNG* gene family (Fig. 1A, lane 6) (19,20), but not enzymes belonging to other classes of uracil-DNA glycosylases such as those represented by *E. coli* Mug (9), *H. sapiens* SMUG1 (21) and *H. sapiens* MBD4 (22). This suggested that uracil processing in the *P. aerophilum* crude extract could be accounted for by at least two distinct type of DNA glycosylase, a major activity that is not inhibited by UGI, and a minor activity that is sensitive to this peptide. However, homology searches of the *P. aerophilum* genome failed to detect genes encoding any UNG-type proteins.

In order to test the effect of temperature on uracil removal from either the double-stranded G•U substrate or from single-stranded DNA, the nicking assay was repeated at different incubation temperatures. The uracil-processing activities present in the *P. aerophilum* extracts were stable at temperatures up to 90°C (Fig. 2), which was anticipated for enzymes of a hyperthermophile. Interestingly, while the processing efficiency of the single-stranded substrate was largely unaffected by incubation temperature, the 60-mer G•U substrate was processed substantially more efficiently at 70°C than at lower temperatures. This suggests that the enzymatic activity increases with temperature, with the optimum lying between 70°C and 90°C under our assay conditions. The activity increase seen at 70°C was not due to melting of the oligonucleotide duplex; first, this would imply a further improvement at 90°C, and second, the single-stranded substrate was processed less efficiently at 70°C than the G•U heteroduplex. However, melting was probably responsible for the observed activity decrease at 90°C (Fig. 2, lane 4), as the processing of the G•U substrate decreased to the level seen with the single-stranded oligonucleotide (Fig. 2, lane 8).

#### *G•T mismatch processing in P. aerophilum extracts*

At 100°C, deamination of 5-methylcytosines is accelerated in parallel with that of cytosines (1). While the latter process generates G•U mispairs in double-stranded DNA, the former reaction gives rise to G•T mismatches. Several years ago, we described a mismatch-specific thymine/uracil DNA-glycosylase, TDG, which removes thymine and uracil from G•T and G•U mismatches (8). We were subsequently able to identify and characterize a bacterial homologue of TDG, named Mug, which removes uracil from G•U mispairs but which has no G•T processing ability (9). At about the same time, Fritz and colleagues identified a similar activity in the thermophile *M. thermoautotrophicum*, named *Mth.MIG* (23), which also processes G•U and

G•T mispairs in a manner analogous to TDG, but which belongs to the endonuclease III family of DNA glycosylases. Notably, the *MIG* gene homologue was recently identified also in *P. aerophilum* by Miller and colleagues and the recombinant protein expressed in *E. coli* was indeed shown to be a mismatch-specific thymine/uracil DNA-glycosylase (14). We decided to test whether the *Pa*-MIG activity could be detected in the *P. aerophilum* crude extracts. Figure 3 (lanes 4 and 6) documents that purified recombinant *Pa*-MIG enzyme removes uracil and thymine mismatched with guanine with about equal efficiency. In contrast, no G•T processing activity could be detected in the crude cell extracts (Fig. 3, lane 5), even when the sensitivity of the assay was increased by the use of <sup>32</sup>P-labelled 90-mer substrates (data not shown). These results, together with the fact that *Pa*-MIG enzyme does not remove uracil residues from A•U pairs or from single stranded DNA, imply that this enzyme does not significantly contribute towards the efficient uracil-processing activity observed in the *P. aerophilum* extracts.

#### *Identification of a P. aerophilum uracil-DNA glycosylase (Pa-UDG)*

During the course of these biochemical studies, the complete sequence of the *P. aerophilum* genome became available (24). We searched this genome database for open reading frames encoding homologues of the UNG or the Mug types of uracil DNA glycosylases, but failed to find likely candidates. This situation changed following the identification of a uracil-DNA glycosylase in the thermophilic eubacterium *T. maritima*, *Tm*-UDG (12). This enzyme was shown to be distantly related to Mug (9), but displayed a substrate specificity similar to that detected in the *P. aerophilum* extracts. We therefore carried out a homology search using the amino acid sequence of *Tm*-UDG. This yielded an ORF encoding a polypeptide with significant homology to *Tm*-UDG (12) and *Af*-UDG (13) (Fig. 4A). A putative translation start was assigned from the genomic sequence and appears to be a GUG rather than an AUG. This is, however, plausible, because GUG is known to be used frequently (up to 25%) as a start codon of archaeal genes (25). The resulting 196-amino acid protein has a calculated molecular mass of 21563 Daltons and a theoretical pI of 9.39. A clustered relationship analysis with UNG, TDG, SMUG and thermophilic uracil DNA glycosylases revealed the existence of four distinct groups of closely-related proteins, where *Pa*-UDG falls into the category of the thermophilic enzymes (Fig. 4B).

#### *Molecular cloning and expression of the P. aerophilum UDG coding sequence; purification of recombinant Pa-UDG enzyme*

The putative *paudg* sequence (PAE0651) was amplified from a genomic clone by high fidelity PCR. The PCR primers used contained restriction sites for *Nhe*I and *Bam*HI, which were then utilized in cloning of the amplified fragment into the corresponding sites within the polylinker of the expression vector pET28c(+). The putative start codon GTG was omitted in the upstream primer, because a suitable ATG translation start site was provided in the expression vector

sequence. The amino terminal of the recombinant protein thus acquired a 6-histidine-tag, as well as a thrombin cleavage site, should the tag need to be removed at a later stage (see Experimental Procedures). The pET28c(+)-*paudg* construct was transformed into competent *E. coli* BL21(DE3) cells and the expression of *paudg* was induced with IPTG. Crude lysates of the IPTG-induced cells contained a protein migrating with an apparent molecular size of ~25 kDa in SDS-PAGE, which was absent from extracts of uninduced cells (data not shown). This is consistent with the predicted MW of a recombinant His<sup>6</sup>-tagged *Pa*-UDG of 24 kDa. The overexpressed protein was purified from the *E. coli* extract using a two-step fractionation scheme involving metal affinity chromatography (nickel-nitrilotriacetic acid) and FPLC (HiTrap SP) (Fig. 5). Starting from one liter of induced *E. coli* culture, we typically obtained ~4 mg of nearly homogeneous recombinant *Pa*-UDG protein at a concentration of ~1 mg/ml. The fraction shown in lane 3 of Figure 5 was used in the subsequent biochemical assays.

*Pa*-UDG is a thermostable, monofunctional uracil DNA-glycosylase that is resistant to UGI

Monofunctional DNA glycosylases generate apurinic or apyrimidinic (AP) sites in DNA, but are unable to convert these into strand breaks, as they lack an associated DNA-lyase activity. Thus, under the conditions of the *in vitro* nicking assay used in this study, the oligonucleotide substrates had to be cleaved at the resulting AP-sites either by the addition of an AP-endonuclease such as HAP1 (26), or by hot alkaline treatment. In contrast, bifunctional DNA glycosylase/lyases can carry out both the base removal and sugar-phosphate backbone cleavage reactions (27). It is impossible to differentiate between these two enzyme classes in assays using crude cell extracts, because these contain also AP-endonuclease activities that can cleave the substrate DNA at the AP-sites. In order to be able to assign *Pa*-UDG to its appropriate class, we carried out the *in vitro* nicking reaction with the purified recombinant enzyme either in the presence or in the absence of HAP1 or NaOH. As shown in Figure 6A, incubation of the 60-mer G•U substrate with the enzyme gave rise to only a small amount of the 23-mer product (lane 4), which arose as a result of spontaneous  $\beta$ -elimination at the AP-site under the conditions of the assay. In contrast, when the oligonucleotide duplexes were subjected to HAP1 or hot alkaline treatment after the incubation with *Pa*-UDG, most of the 60-mer substrate was converted to a 23-mer product, having either a hydroxy group at the 3' end in the case of HAP1 digestion (lane 5) or a phosphate group due to a  $\beta,\delta$ -elimination reaction in the case of NaOH treatment (lane 6). This showed that *Pa*-UDG generates AP-sites in the substrate DNA, which can be cleaved in a subsequent reaction with HAP1 or NaOH. *Pa*-UDG is thus a monofunctional uracil DNA-glycosylase. The recombinant protein did not lose activity when pre-incubated without substrate at 85°C for 15 minutes. This eliminates the possibility that the uracil processing observed in these assays was due to a contamination with UDG from *E. coli*, since the latter activity was reduced to undetectable levels following such heat treatment (data

not shown). Moreover, fractions eluted with 300 mM imidazole from a Ni-NTA column that was loaded with extracts of *E. coli* transfected with the empty pET28c(+) plasmid contained no uracil processing activity (data not shown).

We further examined whether the recombinant *Pa*-UDG is inhibited by UGI. Figure 6B shows that G•U processing by *Pa*-UDG is not affected by the presence of up to 10-fold excess of the inhibitor peptide in the reaction, while an equivalent activity provided by the *E. coli* Ung protein was completely inhibited under similar conditions. We therefore conclude that *Pa*-UDG cannot account for the minor UGI-sensitive uracil DNA glycosylase activity discernible in the *P. aerophilum* extracts.

#### *Substrate specificity of the purified recombinant Pa-UDG*

We next decided to test whether the substrate specificity of the purified recombinant *Pa*-UDG enzyme corresponded to that observed in the crude *P. aerophilum* extracts. Figure 7A shows that this is indeed the case: the enzyme removes uracil from G•U, A•U and single-stranded DNA substrates with an order of preference comparable to that seen in the crude extracts (Fig. 1B). Kinetic time course experiments, in which 1 pmol of G•U, ssU or A•U substrates was incubated with 100 fmol of *Pa*-UDG at 70°C clearly established that the enzyme acts more efficiently on the G•U substrate than on uracil-containing single-stranded DNA or on the A•U duplex (Fig. 7B). 80% of the G•U substrate was cleaved after 30 minutes of incubation, indicating an at least 8-fold enzymatic turnover in this reaction. The activity on the ssU and A•U substrates was lower, with only ~30% being processed after 30 minutes under similar assay conditions. No activity was discernible with the G•T substrate (not shown). The observed turnover on the G•U mismatch is in line with the biochemical properties of UNG-type uracil DNA glycosylases, but contrasts with those of TDG and Mug enzymes that were shown to be strongly product-inhibited (9,28,29). Correspondingly, the latter enzymes bind to oligonucleotide substrates containing the product AP-site in electrophoretic mobility shift assays (EMSA) with high efficiency. *Pa*-UDG failed to form a protein/DNA complex detectable in EMSA with any of the substrates tested (data not shown).

Taken together, the resistance of purified recombinant *Pa*-UDG to the UGI peptide, its ability to turn over in *in vitro* assays and its preference for G•U substrate over A•U and uracil in single-stranded DNA leads to the conclusion that this enzyme is different from UNG or Mug type uracil DNA glycosylases and that is most likely responsible for the major uracil processing activity observed in the crude *P. aerophilum* extracts.

## DISCUSSION

The extracts of *P. aerophilum* contain an efficient uracil-processing activity. We identified and expressed in *E. coli* a *P. aerophilum* ORF that encodes a putative homologue of thermophilic uracil DNA glycosylases, *Pa*-UDG. We showed that the recombinant protein acts as a monofunctional uracil DNA-glycosylase that can remove the aberrant base from single stranded DNA, as well as from A•U pairs and G•U mispairs. As the substrate specificity of this enzyme corresponded to that seen in the total extracts of *P. aerophilum*, we propose that this enzyme accounts for the major uracil processing activity in the extracts of the hyperthermophilic archaeon. In contrast, the *Pa*-MIG activity, which removes thymine and uracil residues from mispairs with guanine (14) could not be detected in the extracts of this organism.

*Pa*-UDG is most closely related to two recently identified thermostable uracil DNA glycosylases from *T. maritima* and *A. fulgidus* (12,13), and these three proteins branch off as a separate cluster of uracil DNA glycosylases in phylogenetic analysis (Fig. 4B). This distinction is not only evident at the level of sequence similarity, but also by their biochemical properties such as substrate specificity, stability of protein/DNA interactions and enzymatic turnover. However, the amino acid sequences around the putative active sites of the different uracil glycosylases are interesting and may help us understand more about the evolution and function of these proteins (see also references (30,31)). Crystal structures of the UNG-type uracil glycosylases from *Herpes simplex* virus (32) and *Homo sapiens* (33), as well as that of the *E. coli* Mug protein (34,35), have shown that the enzymes have a structurally conserved, bipartite active site (Fig. 8). Motif A in the UNG-type enzymes contains the sequence GQDPYH...F, where the aspartic acid residue (shown bold in a gray field) activates a molecule of water towards a nucleophilic attack on the C1' of the sugar residue bound to the aberrant base. The tyrosine ensures exclusion of thymine from the active site pocket, while the phenylalanine in the second conserved motif stabilizes uracil in the pocket *via*  $\pi$ - $\pi$  interactions. The histidine residue of the conserved B-motif HPSPLS is thought to activate the glycosidic bond towards nucleophilic attack by the water molecule through binding to the O<sub>2</sub> of the uracil (32). In the Mug structure (34), both the aspartate and the histidine are replaced with asparagines, which can still position the water molecule correctly for the attack, but cannot activate it as efficiently. This was thought to be the underlying reason for the low enzymatic activity of Mug and TDG as compared to the UNG-type UDG enzymes (34,35). However, the issue must be somewhat more complex, as substitution of the TDG motif A (GINPG) with GQDPG, and motif B (MPS) with HPS gave rise to an inactive enzyme (U. Hardeland, J. Jiricny and P. Schär, unpublished).

Within the group of thermostable UDGs, the A motif appears to be a hybrid of the UNG- and Mug-type enzymes. Upon first examination, it resembles more closely that of the Mug class, at least as far as the second conserved glycine residue of the motif GxxPG is



concerned. However, unlike the Mug family, the A motif of the thermophiles has an acidic residue (E), which may well assist in the activation of the catalytic water molecule, as suggested for the UNG-type enzymes (30,31). The tyrosine residue in the GQDPY motif of the UNG-type UDGs was proposed to prevent the binding of thymine in the active site pocket (32,33) and it may be assumed that substitution of this residue for the small glycine would allow the processing of this base. While this may be true for UNG-type UDG's (36), of all the enzymes with the GxxPG motif (Fig. 8) only the human TDG removes thymines from G•T mispairs (9), which suggests that the thymine exclusion criteria are more complex than previously thought. The A-motif of the hyperthermophiles also contains several acidic residues (shown in bold in Fig. 8), which are absent from this region of the other proteins. Their biological role may lie in the stabilization of the active site through salt bridges.

Analysis of the B-motif is also interesting. The thermophilic UDGs, as well as the UNG- and SMUG-type enzymes, share the conserved sequence HPS (or HPA in the archaea), while the histidine residue is absent from the corresponding site of the Mug enzymes (Fig. 8). Comparison of the substrate specificities of the HPS/A group revealed that they are all able to process single-stranded DNA substrates, while Mug and TDG cannot. However, the Mug homologue from *S. pombe* has the sequence GIS, thus lacking both the histidine and the proline residues, yet it can remove uracil from ssDNA (U. Hardeland and P. Schär, unpublished). It thus appears unlikely that the histidine in this motif is required for ssDNA processing. Interestingly, following removal of the uracil, the Mug enzymes remain bound to the AP-site containing oligonucleotides (9,29) and hence fail to turn over. In contrast, all the HPS/A enzymes, with the notable exception of SMUG, display turnover kinetics. The crystal structure of substrate-bound Mug revealed that, similarly to the UNG-type proteins, this enzyme inserts a wedge of three amino acids into the DNA duplex in order to fill the space vacated by the flipped-out uracil, thus preventing the neighboring bases from collapsing onto each other. The structural organization of this wedge involves the residues of Mug motif B. Unlike UNG, however, Mug establishes specific hydrogen bonds with the Watson-Crick face of the widowed guanine opposite the abasic site, as well as with the sugar residues of the complementary strand (34). Since the UNG- (32,33) and the thermophilic-type UDG's do not appear to engage in rigid complementary strand interactions, it may be this property of Mug-type enzymes and thus this function of motif B that is responsible for the observed differences in turnover kinetics. This hypothesis would appear to be supported by the finding that a MPS to HPS mutation within the motif B of the human TDG generated an active enzyme that has lost its ability to stably interact with the DNA substrate (29). However, though it may be an important contributing factor, it is unlikely that a single amino acid sequence motif is entirely responsible for controlling the dissociation of the enzyme from its substrate and further studies are required in order to elucidate these differences.



The genome of *P. aerophilum* (24) contains two ORFs encoding conserved uracil glycosylases: *Pa*-MIG (14) and *Pa*-UDG. Both these enzymes could be expressed in *E. coli* and subjected to biochemical analysis, which revealed that they were not inhibited by the generic uracil glycosylase inhibitor UGI. As similar experiments carried out with crude extracts of *P. aerophilum* revealed a low degree of inhibition of uracil processing, we conclude that they must contain yet another enzyme, capable of addressing this pre-mutagenic product of hydrolytic deamination of cytosine. Homology sequence searches failed to find a third ORF that would contain the two amino acid motifs thought to comprise the active site of the UDG family of enzymes. We are currently attempting to identify this elusive activity by biochemical means.

### Acknowledgements

The authors wish to express their gratitude to Mahmud Shivji for the *P. aerophilum* extracts, to Hanjin Yang for the gift of the purified *Pa*-MIG, to Ulrike Hardeland for assistance and for the gift of the purified human HAP1 endonuclease and to Orlando Schärer for critical reading of the manuscript. The work was supported by grants from the UBS to A. S. and from the Swiss National Science Foundation to P.S.

### REFERENCES

1. Lindahl, T. and Nyberg, B. (1974) *Biochemistry* **13**, 3405-3410.
2. Jacobs, K. L. and Grogan, D. W. (1997) *J. Bacteriol.* **179**, 3298-3303.
3. Koulis, A., Cowan, D. A., Pearl, L. H. and Savva, R. (1996) *FEMS Microbiol. Lett.* **143**, 267-271.
4. Volkl, P., Huber, R., Drobner, E., Rachel, R., Burggraf, S., Trincone, A. and Stetter, K. O. (1993) *Appl. Environ. Microbiol.* **59**, 2918-2926.
5. Krokan, H. E., Standal, R. and Slupphaug, G. (1997) *Biochem. J.* **325**, 1-16.
6. Wiebauer, K. and Jiricny, J. (1989) *Nature* **339**, 234-236.
7. Neddermann, P., Gallinari, P., Lettieri, T., Schmid, D., Truong, O., Hsuan, J. J., Wiebauer, K. and Jiricny, J. (1996) *J. Biol. Chem.* **271**, 12767-12774.
8. Neddermann, P. and Jiricny, J. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 1642-1646.
9. Gallinari, P. and Jiricny, J. (1996) *Nature* **383**, 735-738.
10. Aravind, L., Walker, D. R. and Koonin, E. V. (1999) *Nucleic Acids Res.* **27**, 1223-1242.
11. Fitz-Gibbon, S., Choi, A. J., Miller, J. H., Stetter, K. O., Simon, M. I., Swanson, R. and Kim, U. J. (1997) *Extremophiles* **1**, 36-51.
12. Sandigursky, M., and Franklin, W. A. (1999) *Curr. Biol.* **9**, 531-534.
13. Sandigursky, M., and Franklin, W. A. (2000) *J. Biol. Chem.*, **275**, 19146-19149.

14. Yang, H., Fitz-Gibbon, S., Marcotte, E. M., Tai, J. H., Hyman, E. C. and Miller, J. H. (2000) *J. Bacteriol.* **182**, 1272-1279.
15. Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory Press.
16. Hudepohl, U., Reiter, W. D. and Zillig, W. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 5851-5855.
17. Devereux, J., Haeberli, P. and Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387-395.
18. Bradford, M. M. (1976) *Anal. Biochem.* **72**, 248-254.
19. Wang, Z. and Mosbaugh, D. W. (1989) *J. Biol. Chem.* **264**, 1163-1171.
20. Wang, Z. G., Smith, D. G. and Mosbaugh, D. W. (1991) *Gene* **99**, 31-37.
21. Haushalter, K. A., Todd Stukenberg, M. W., Kirschner, M. W. and Verdine, G. L. (1999) *Curr. Biol.* **9**, 174-185.
22. Hendrich, B., Hardeland, U., Ng, H. H., Jiricny, J. and Bird, A. (1999) *Nature* **401**, 301-304.
23. Horst, J. P. and Fritz, H. J. (1996) *EMBO J.* **15**, 5459-5469.
24. Fitz-Gibbon, S.T., Ladner, H., Kim, U.-J., Stetter, K.O., Simon, M.I. and Miller, J. H. (submitted).
25. Dennis, P. P. (1997) *Cell* **89**, 1007-1010.
26. Robson, C. N. and Hickson, I. D. (1991) *Nucleic Acids Res.* **19**, 5519-5523.
27. Dodson, M. L., Michaels, M. L. and Lloyd, R. S. (1994) *J. Biol. Chem.* **269**, 32709-32712.
28. Waters, T. R., Gallinari, P., Jiricny, J. and Swann, P. F. (1999) *J. Biol. Chem.* **274**, 67-74.
29. Hardeland, U., Bentele, M., Jiricny, J. and Schar, P. (2000) *J. Biol. Chem.* **275**, 33449-33456.
30. Pearl, L. H. (2000) *Mutat. Res.* **460**, 165-181.
31. Aravind, L. and Koonin, E. V. (2000) *Genome Biol.* **1**, 1-8.
32. Savva, R., McAuley-Hecht, K., Brown, T. and Pearl, L. (1995) *Nature* **373**, 487-493.
33. Mol, C. D., Arvai, A. S., Slupphaug, G., Kavli, B., Alseth, I., Krokan, H. E. and Tainer, J. A. (1995) *Cell* **80**, 869-878.
34. Barrett, T. E., Savva, R., Panayotou, G., Barlow, T., Brown, T., Jiricny, J. and Pearl, L. H. (1998) *Cell* **92**, 117-129.
35. Barrett, T. E., Scharer, O. D., Savva, R., Brown, T., Jiricny, J., Verdine, G. L., and Pearl, L. H. (1999) *EMBO J.* **18**, 6599-6609.
36. Kavli, B., Slupphaug, G., Mol, C. D., Arvai, A. S., Peterson, S. B., Tainer, J. A. and Krokan, H. E. (1996) *EMBO J.* **15**, 3442-3447.
37. Corpet, F. (1988) *Nucleic Acids Res.* **16**, 10881-10890.

## LEGENDS TO FIGURES

**Figure 1: Uracil DNA glycosylase activities in extracts of *P. aerophilum*.** **A**, G•U mismatch processing in crude extracts of *P. aerophilum* and *E. coli* strain BL21(DE3). The 60-mer substrates were incubated with 1  $\mu$ g of *P. aerophilum* extract or 1.5  $\mu$ g of *E. coli* extract for 1 h at 37°C in the presence or absence of 10 Units of the peptide inhibitor UGI as described in ‘Experimental Procedures’. *Lane 1*, G•C 60-mer incubated with *P. aerophilum* extract; *lane 2*, untreated G•U 60-mer; *lane 3*, G•U 60-mer incubated with *P. aerophilum* extract; *lane 4*, G•U 60-mer incubated with *P. aerophilum* extract pre-treated with 10 U of UGI for 5 min at 37°C; *lane 5*, G•U 60-mer incubated with *E. coli* extract; *lane 6*, G•U 60-mer incubated with *E. coli* extract pre-treated with 10 U of UGI for 5 min at 37°C. **B**, Substrate specificity of uracil-processing activities in extracts of *P. aerophilum*. The 60-mer substrates were incubated with 1  $\mu$ g of *P. aerophilum* extract for 1 h at 45°C as described in ‘Experimental Procedures’. *Lane 1*, 23-mer marker produced by digestion of the G•C 60-mer with *AccI*; *lanes 2, 4 and 6*, untreated 60-mer substrates G•U, A•U and ssU, respectively; *lanes 3, 5 and 7*, same 60-mers incubated with extract. The positions of the full length substrate and of the 23-mer product are indicated. The fluorescently-labeled DNA strand in the substrate is indicated by an asterisk.

**Figure 2: Effect of temperature on uracil processing in *P. aerophilum* extracts.** The 60-mer oligonucleotides were incubated with 1  $\mu$ g of extract for 15 min at the indicated temperatures. The extract was heat-treated for 15 min at 80°C prior to addition to the substrate. *Lanes 1-4*, G•U 60-mer; *lanes 5-8*, single-stranded uracil-containing 60-mer. The fluorescently-labeled DNA strand is indicated by an asterisk.

**Figure 3: G•U and G•T mismatch processing by crude extracts of *P. aerophilum* and by the purified recombinant *Pa*-MIG protein.** The DNA substrates were incubated with either 25  $\mu$ g of *P. aerophilum* extract or with 10 pmol of *Pa*-MIG for 1 h at 45°C. Due to the lack of endonuclease activity in the purified enzyme preparation, the reactions with *Pa*-MIG were stopped by the addition of NaOH to 100 mM and by heating for 10 min at 99°C. The 23-mer product bands in lanes 4 and 6 thus carry 3'-phosphate groups and migrate faster through polyacrylamide gels than the product in the extract reaction (lane 3), which arose through cleavage of the AP-site by a hydrolytic AP-endonuclease present in the crude extracts that generates fragments carrying 3'-OH groups. The fluorescently-labeled strand is indicated by an asterisk.

**Figure 4: Sequence conservation and phylogenetic categorization of the *P.***

**aerophilum uracil-DNA glycosylase *Pa*-UDG.** **A,** Amino acid sequence alignment of *P. aerophilum* uracil-DNA glycosylase with homologues from *A. fulgidus* and *T. maritima*. Identical residues are shaded and the two amino acid motifs (A and B) contributing to the folds of the putative active site pockets of the enzymes are underlined. The *Pa*-UDG ORF was identified in the sequence of the *P. aerophilum* genome by using TFASTA (gap creation penalty 3.0, gap extension penalty 0.1) of the Genetics Computer Group (GCG) software package (Version 10.0). The sequence alignment shown was performed using the MultAlin software (37) available at <http://www.toulouse.inra.fr>; **B,** Phylogenetic tree representing the clustered relationship between different variants of uracil DNA glycosylases. The amino acid sequence alignment underlying the phylogenetic output was generated with the PILEUP program (gap creation penalty 5.0, gap extension penalty 0.1) of the GCG software package (Version 10.0). The proteins shown are: spTHP1, *S. pombe* TDG homologue Thp1 (EMBL: AJ277958); ecMUG, *E. coli* TDG homologue Mug (Swissprot: P43342); hsTDG, *H. sapiens* TDG (EMBL: U51166); ecUDG, *E. coli* uracil DNA glycosylase Ung (Swissprot: P12295); hsvUDG, *herpes simplex* virus 1 uracil DNA glycosylase (Swissprot: P10186); hsUDG, *H. sapiens* uracil DNA glycosylase UNG1 (Swissprot: P13051); spUDG, *S. pombe* uracil DNA glycosylase UDG (Swissprot: O74834); paUDG, *P. aerophilum* uracil DNA glycosylase (this study); afUDG, *A. fulgidus* uracil DNA glycosylase (13); tmUGD, *T. maritima* uracil DNA glycosylase (12); hsSMUG1, *H. sapiens* single strand specific uracil DNA glycosylase SMUG1 (EMBL: AF125182); xISMUG1, *X. laevis* single strand specific uracil DNA glycosylase SMUG1 (Swissprot: Q9YGN6); dmSMUG, putative *D. melanogaster* single-strand specific uracil DNA glycosylase (ENTREZ: AAF55400)

**Figure 5: Purification of recombinant *Pa*-UDG expressed in *E. coli*.** *Lane 1*, *Pa*-UDG-containing eluate from Ni-NTA after dialysis; *lane 2*, HiTrap SP FPLC column, flow-through fraction; *lane 3*, HiTrap SP FPLC column, eluted *Pa*-UDG protein fraction after dialysis; *M*, broad range molecular mass standards (Bio-Rad) as indicated on the left. The band due to the recombinant *Pa*-UDG is indicated. The figure shows a Coomassie Blue stained SDS-polyacrylamide gel (15%).

**Figure 6: *Pa*-UDG is a monofunctional uracil DNA glycosylase that is resistant to the specific UDG inhibitor UGI.** **A,** The 60-mer substrates were incubated with 0 (lanes 2 and 3) or 0.5 pmol (lanes 4-6) of *Pa*-UDG for 30 min at 37°C. *Lane 1*, G•C substrate cleaved with *AccI* at 37°C for 30 min; *lane 2*, G•U substrate digested with 35µM HAP1 at 37°C for 5min; *lane 3*, G•U substrate treated with 100 mM NaOH at 99°C for 10 min; *lane 4*, G•U substrate incubated with 0.5 pmol of *Pa*-UDG; *lane 5*, G•U substrate incubated with *Pa*-UDG

followed by digestion with 35  $\mu$ M HAP1 at 37°C for 5 min; *lane 6*, G•U substrate incubated with *Pa*-UDG, followed by treatment with 100 mM NaOH at 99°C for 10 min. The fluorescently labeled DNA strand is indicated by an asterisk. The faint product band in *lane 4* is due to spontaneous  $\beta$ -elimination of the labile AP-sites. **B**, The effect of UGI on G•U mismatch processing by purified recombinant *Pa*-UDG. The G•U 60-mer substrates were incubated with 100 fmol of *Pa*-UDG or 2 Units of *Ec*-UDG for 1 h at 37° as described in ‘Experimental Procedures’. The reactions were stopped by the addition of NaOH to 100 mM, followed by heating at 99°C for 10 min. *Lane 1*, untreated G•U 60-mer substrate; *lane 2*, G•U substrate incubated with *Pa*-UDG; *lanes 3 and 4*, G•U substrates incubated with *Pa*-UDG pre-treated for 5 min at 37°C with 1 or 10 U of UGI, respectively; *lane 5*, G•U substrate incubated with *Ec*-UDG; *lanes 6 and 7*, G•U substrates incubated for 5 min at 37°C with *Ec*-UDG pre-treated with 1 or 10 U of UGI, respectively. The fluorescently- labeled DNA strand is indicated by an asterisk.

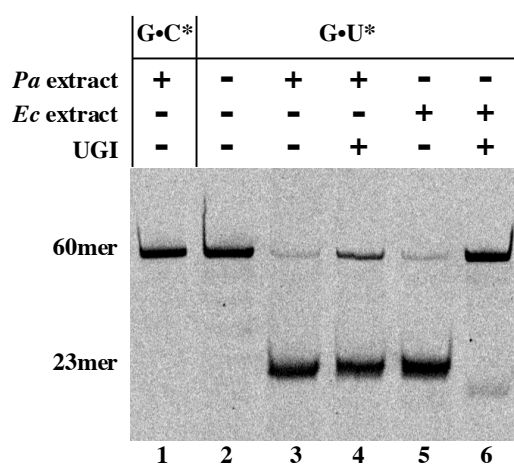
**Figure 7: Substrate specificity of *Pa*-UDG.** **A**, 100 fmol of *Pa*-UDG were incubated with the G•C, G•U, A•U or the single-stranded U 60-mer oligonucleotides at 37°C for 1 h. The reactions were stopped by the addition of NaOH to 100 mM and by heating at 99°C for 10 min. *Lanes 1, 2, 4 and 6*, substrates incubated with *Pa*-UDG; *lanes 3, 5 and 7*, substrates incubated with *Pa*-UDG in the presence of 1 U of UGI peptide. The fluorescently-labeled strand is indicated by an asterisk. **B**, Time course of the *Pa*-UDG catalyzed removal of uracil from G•U (●) and A•U (■) 60-mer substrates. The substrate oligonucleotides were incubated with 100 fmol of *Pa*-UDG for 1, 3, 5, 10, 15 and 30 min at 70°C. The reactions were stopped by the addition of NaOH to 100 mM and by heating at 99°C for 10 min. The amounts of the 23-mer product were quantitated using the ImageQuant software. The processing kinetics of the ssU substrate were indistinguishable from those of A•U.

**Figure 8: Evolutionary conservation of the catalytic domain of uracil DNA glycosylases.**

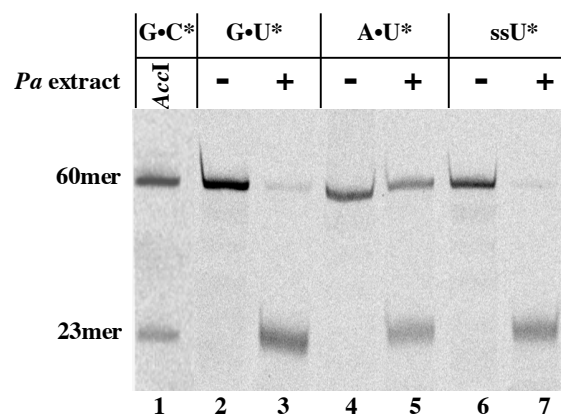
Shown are partial amino acid sequence alignments of the active site motifs A and B (see also Fig. 4a) of the UNG-type uracil DNA glycosylases from *H. sapiens* (hsUDG, Swissprot: P13051) and human *herpes simplex virus 1* (hsvUDG, Swissprot: P10186), the thermophilic UDGs from *P. aerophilum*, *A. fulgidus* (13) and *T. maritima* (12), SMUG1 from *H. sapiens* (EMBL: AF125182) and the MUG/TDG proteins from *E. coli* (Swissprot: P43342), *S. pombe* (EMBL: AJ277958) and *H. sapiens* (EMBL: U51166). Highly-conserved residues are shown in black boxes, with the exception of two conservative changes that are highlighted in dark

gray. The putative residues required for catalytic activity in the A motif are marked by gray boxes and additional acidic residues in the A motif of the hyperthermophiles are shown in bold.

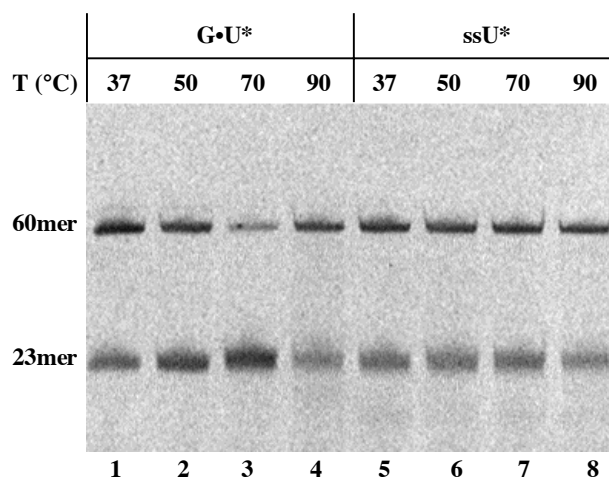
A

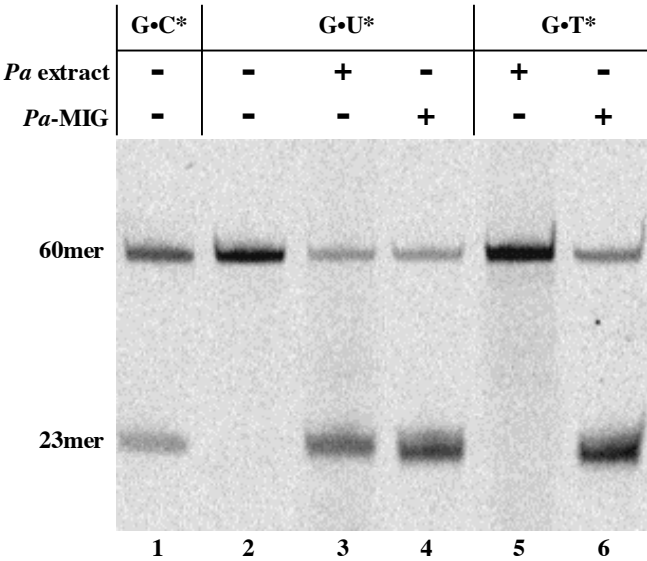


B





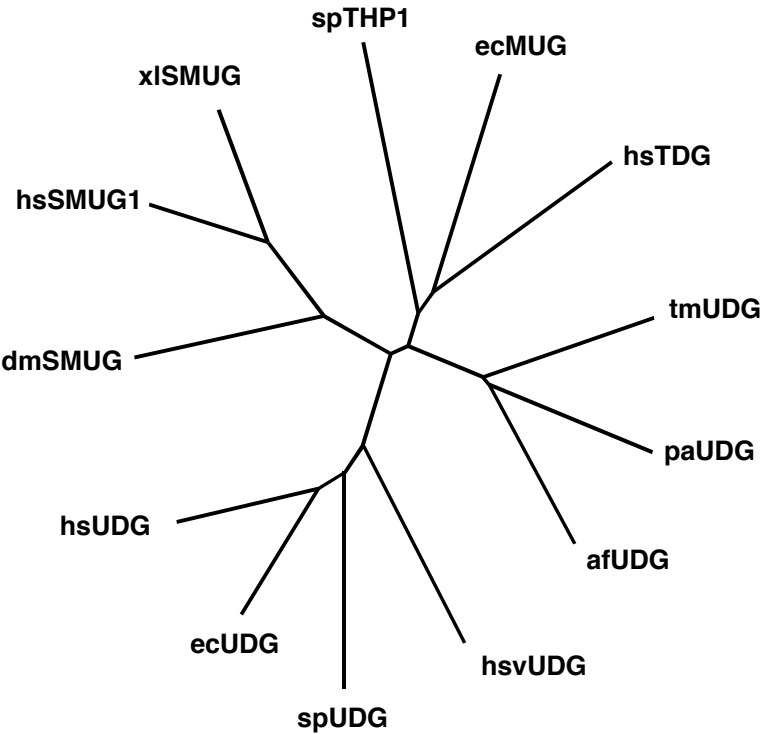


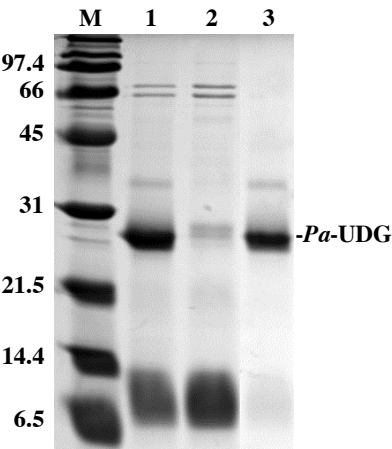


A

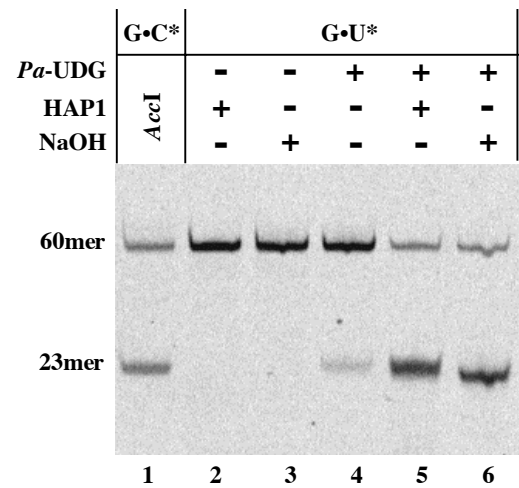
<i>P. aerophilum</i>	1	-	M	D	L	Q	K	L	H	E	L	I	K	N	C	D	K	C	P	L	H	K	Y	R	K	N	A	V	P	G	E	G	E	M	K	L	34
<i>A. fulgidus</i>	1	M	E	S	L	D	D	I	V	R	E	I	M	S	C	R	K	C	D	L	H	K	T	K	T	N	Y	V	P	G	V	G	N	E	K	A	35
<i>T. maritima</i>	1	-	-	-	M	E	I	V	S	E	R	V	K	K	C	T	A	C	P	L	H	L	N	R	T	N	V	V	V	G	E	G	N	L	D	T	32
<i>P. aerophilum</i>	35	G	V	M	I	V	G	E	A	P	G	A	S	E	D	E	A	G	R	P	F	V	G	A	A	G	Q	L	L	T	E	A	L	S	R	L	69
<i>A. fulgidus</i>	36	E	I	V	F	V	G	E	A	P	G	R	D	E	D	L	K	G	E	P	F	V	G	A	A	G	K	L	L	T	E	M	L	A	S	I	70
<i>T. maritima</i>	33	R	I	V	F	V	G	E	G	P	G	E	E	D	K	T	G	R	P	F	V	G	R	A	G	M	L	L	T	E	L	L	R	E	S	67	
A																																					
<i>P. aerophilum</i>	70	G	V	R	R	G	D	V	F	I	T	N	V	V	K	C	R	P	P	N	N	R	T	P	N	R	E	E	V	E	A	C	L	P	Y	L	104
<i>A. fulgidus</i>	71	G	L	R	R	E	D	V	Y	I	T	N	V	L	K	C	R	P	P	N	N	R	D	P	T	P	E	E	V	E	K	C	G	D	Y	L	105
<i>T. maritima</i>	68	G	I	R	R	E	D	V	Y	I	C	N	V	V	K	C	R	P	P	N	N	R	T	P	T	P	E	E	Q	A	A	C	G	H	F	L	102
<i>P. aerophilum</i>	105	I	Q	Q	I	G	I	L	K	P	R	R	I	I	A	L	G	L	I	S	A	K	A	L	M	E	L	M	G	R	R	A	E	K	L	G	139
<i>A. fulgidus</i>	106	V	R	Q	L	E	A	I	R	P	N	V	I	V	C	L	G	R	F	A	A	Q	F	I	F	N	L	F	D	L	E	F	T	T	I	S	140
<i>T. maritima</i>	103	L	A	Q	I	E	I	I	N	P	D	V	I	V	A	L	G	A	T	A	L	S	F	F	V	D	-	-	-	G	K	V	S	I	T	134	
<i>P. aerophilum</i>	140	D	V	K	G	K	C	Y	Q	G	R	I	A	G	V	Q	V	E	L	C	I	T	Y	H	P	A	A	V	L	R	K	P	A	-	-	L	172
<i>A. fulgidus</i>	141	R	V	K	G	K	V	Y	E	V	E	R	W	G	K	K	V	K	V	I	A	I	Y	H	P	A	A	V	L	Y	R	P	Q	-	-	L	173
<i>T. maritima</i>	135	K	V	R	G	-	-	-	N	P	I	D	W	L	G	G	K	K	V	I	P	T	F	H	P	S	Y	L	L	R	N	R	S	N	E	L	166
B																																					
<i>P. aerophilum</i>	173	R	G	E	F	Q	K	D	L	A	M	F	F	G	G	G	L	D	R	F	L	D	P	S	K	-	-	-	-	-	-	-	-	-	-	196	
<i>A. fulgidus</i>	174	R	E	E	Y	E	S	D	F	K	K	I	G	E	L	C	G	K	K	Q	P	T	L	F	D	Y	L	-	-	-	-	-	-	-	-	199	
<i>T. maritima</i>	167	R	R	I	V	L	E	D	I	E	K	A	K	S	F	I	K	K	E	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	185	

B

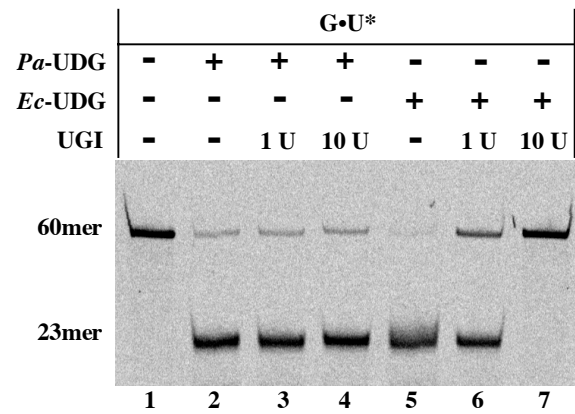




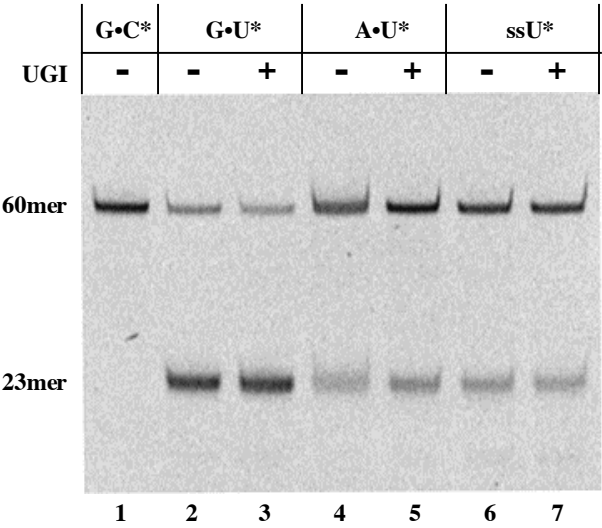
A



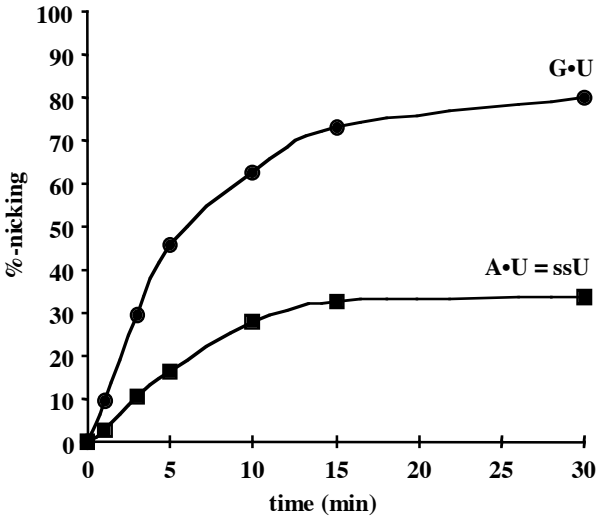
B



A



B



<i>Hs</i> -UDG	143	<b>G</b>	Q	<b>D</b>	<b>P</b>	Y	H	G	P	N	Q	A	H	<b>G</b>	L	C	<b>F</b>	158	268	<b>H</b>	<b>P</b>	<b>S</b>	P	L	S	273
<i>hsv</i> UDG	86	<b>G</b>	Q	<b>D</b>	<b>P</b>	Y	H	H	P	G	Q	A	H	<b>G</b>	L	A	<b>F</b>	101	210	<b>H</b>	<b>P</b>	<b>S</b>	P	L	S	215
<i>Pa</i> -UDG	40	<b>G</b>	E	A	<b>P</b>	<b>G</b>	-	A	S	<b>E</b>	<b>D</b>	E	A	<b>G</b>	R	<b>P</b>	<b>F</b>	54	162	<b>H</b>	<b>P</b>	<b>A</b>	A	V	L	167
<i>Af</i> -UDG	41	<b>G</b>	E	A	<b>P</b>	<b>G</b>	-	R	<b>D</b>	<b>E</b>	<b>D</b>	L	K	<b>G</b>	E	<b>P</b>	<b>F</b>	55	163	<b>H</b>	<b>P</b>	<b>A</b>	A	V	L	168
<i>Tm</i> -UDG	38	<b>G</b>	E	G	<b>P</b>	<b>G</b>	-	E	E	<b>E</b>	<b>D</b>	K	T	<b>G</b>	R	<b>P</b>	<b>F</b>	52	154	<b>H</b>	<b>P</b>	<b>S</b>	Y	L	L	159
<i>Hs</i> -SMUG1	83	<b>G</b>	M	<b>N</b>	<b>P</b>	<b>G</b>	P	F	G	M	A	Q	T	<b>G</b>	V	<b>P</b>	<b>F</b>	98	239	<b>H</b>	<b>P</b>	<b>S</b>	P	R	N	244
<i>Ec</i> -MUG	16	<b>G</b>	I	<b>N</b>	<b>P</b>	<b>G</b>	-	L	S	S	A	G	T	<b>G</b>	F	<b>P</b>	<b>F</b>	30	140	N	<b>P</b>	<b>S</b>	G	L	S	145
<i>Sp</i> -Thp1	156	<b>G</b>	L	<b>N</b>	<b>P</b>	<b>G</b>	-	I	T	S	S	L	K	<b>G</b>	H	A	<b>F</b>	170	288	G	I	<b>S</b>	S	S	G	293
<i>Hs</i> -TDG	138	<b>G</b>	I	<b>N</b>	<b>P</b>	<b>G</b>	-	L	M	A	A	Y	K	<b>G</b>	H	H	<b>Y</b>	152	269	M	<b>P</b>	<b>S</b>	S	S	A	274

A

B